ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 12, December 2018

An Implementation of Authenticity of Profiles against Spoofing by Pre-emptive Mechanism on OSNs

Divya Naga Lakshmi Chunduri[1], Dr M. Sampath Kumar[2]

M.Tech Scholar, Department of Computer Science and System Engineering, Andhra University College of Engineering (A), Visakhapatnam, AP, India[1]

Department of Computer Science and System Engineering, Andhra University College of Engineering (A), Visakhapatnam, AP, India[2]

ABSTRACT*:* In many OSN applications, developers collect only a limited sample of tweets and a local portion of the Twitter network. We develop a collection of network-, linguistic-, and application oriented variables that could be used as possible features, and identify specific features that distinguish well between humans and bots. The popularity and open structure of OSN have attracted a large number of automated programs, known as bots, which appear to be a double-edged sword to Twitter. Legitimate bots generate a large amount of benign tweets delivering news and updating feeds, while malicious bots spread spam or malicious contents. More interestingly, in the middle between human and bot, there has emerged cyborg referred to either bot-assisted human or human-assisted bot. To assist human users in identifying who they are interacting with, this paper focuses on the classification of human, bot, and cyborg accounts on Twitter. We first conduct a set of large-scale measurements with a collection of over 500,000 accounts. We observe the difference among human, bot in terms of tweeting behaviour, tweet content, and account properties. Based on the measurement results, we propose a classification system that includes the following four parts: 1) an entropy-based component, 2) a spam detection component, 3) an account properties component, and 4) a decision maker. It uses the combination of features extracted from an unknown user to determine the likelihood of being a human, bot, or cyborg. Our experimental evaluation demonstrates the efficacy of the proposed classification system. The research discussed in this paper applies these same engineered features to a set of fake human accounts in the hope of advancing the successful detection of fake identities created by humans on OSN.

KEYWORDS: Automatic identification, bot , Twitter, social networks, OSN, SMP.

I. INTRODUCTION

The openness of Twitter's platform allows for, and even promotes, programs that automatically post. These "bots" post content ranging from helpful (e.g., recent news stories or public service announcements) to malicious (e.g., spam or phishing links). Such malicious bots on Twitter have become a nuisance, even recently triggering a long diatribe in the New Yorker [1]. They are alleged to help political candidates skew public perception; for instance, botornet.net asserts that former US Presidential candidate Newt Gingrich gained over a million followers on Twitter through the use of bots, a charge Mr. Gingrich is reported to have denied. Similarly, Hill [2] reports that "Facebook thinks 83 million of its users are fake" and that "up to 29.9% of Barack Obama's 17.82 million [Twitter] followers and 21.9% of Mitt Romney's 814,000 followers may be fake". In short, there is now a widespread belief that bots constitute a significant part of the social media world—and that many of them are malicious in their intent. In this paper, we study the problem of identifying bots on Twitter from an application perspective. Few real-world applications analyze all of Twitter. Rather, most applications focus on those portions of the Twitter tweet database and the Twitter network that are relevant to

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

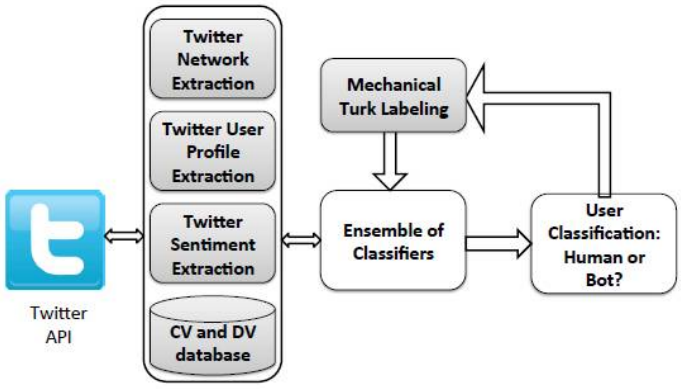
Vol. 7, Issue 12, December 2018

them. For example, if a politician were interested in tracking the recent Indian elections, he would probably identify a set of topics of interests (TOI) consisting of names of relevant politicians and relevant political parties. The identification of bots may then be based on a relevant subset of tweets and a relevant subset of the Twitter network that are TOI focused. Compared to past work [3] on bot identification in Twitter, TOI-focused applications may not have a huge network with which to work, especially as reconstructing a network from Twitter's open API can pose some challenges. In such industrial applications, there is a critical need to identify bots:

1) For instance, in an election application we built, we needed to predict the expected number of supporters for a particular candidate. But in this prediction, we needed to discount for bots.
2) We worked with a company that wanted to identify the most influential Twitter users on a suite of topics of interest to the company. They wanted to be sure that the measure of influence discounted for bots (and of course that bots were not included in the answer).
In contrast to past work, we approach this task from a new viewpoint, namely analyzing the novel semantic feature of tweet sentiment in human and bot accounts. To our knowledge, no prior work has used tweet sentiment to separate human and non-human users on Twitter. We present SentiBot, a sentiment-aware architecture for identifying bots on Twitter. We test it on a real dataset that does not include malicious accounts already suspended by Twitter's fielded bot detection algorithms, but—as validated by our human labelers still includes bots that slipped through their filter. Including sentiment-aware features in the classification process improves accuracy on these "harder" classification cases, where presently fielded algorithms fail.

II. PREVIOUS WORK

There has been recent interest in the detection of malicious and/or fake users from both the online social networks and computer networking communities. For instance, Wang [4] looks at graph-based features to identify bots on Twitter, while Yang, Harkreader, and Gu [5] combine similar graph based features with syntactic metrics to build their classifiers. Thomas et al. [6] use a similar set of features to provide a retrospective analysis of a large set of recently-suspended Twitter accounts. Boshmaf et al. [7] instead create bots (rather than detecting them), claiming that 80% of bots are undetectable and that Facebook's Immune system [8] was unable to detect their bots. Lee, Caverlee, and Webb [9] create "honeypot" accounts to lure both humans and spammers into the open, then provide a statistical analysis of the malicious accounts they identified. In computer networks research, the detection of Sybil accounts in computer networks has been applied to social network data; these techniques tend to rely on the "fast mixing" property of a network—which may not exist in social networks —and do not scale to the size of present-day social networks.



ARCHITECTURE

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 12, December 2018

III. DATASET

We use the "India Election Dataset" (IEDS), a real-world datasetet that was collected from July 15, 2013 to March 24, 2014. A tweet is considered "on topic" (and included in IEDS) if it includes one of a number of keywords pertaining to the election (e.g., Indian political parties like "Shiv Sena" or "BJP", politicians like "Rajnath Singh" or "Nitish Kumar", etc.) or if it was tweeted by a user who frequently tweets about such political keywords. We discuss the set of these "topics of interest" (TOI) later in this section. Critically, there exists a strong incentive to deploy bots in the IEDS dataset, as noted by the Indian press' coverage of the use of bots and fake accounts during this election cycle [20].

Twitter Sentiment Extraction.

Bot used the Twitter API to first identify a set of users during the selected time frame who had tweeted on at least one "topic of interest" in a set TOI. A TOI is any set of terms. In the case of IEDS, the TOI consisted of politicians and political parties involved in the election. For each day d, each user, and each I, we calculated the sentiment score SS(d; u; t) of user u on topic t, averaged across all tweets on topic t posted by the user on that day. We used SentiMetrix's commercially available sentiment scoring engine [17], [18] that assigns a value between 1 ("maximally negative") and +1 ("maximally positive") to score the intensity of sentiment on topic t in a tweet. Of course, this scoring engine can be swapped out with other similar scoring engines such as those due to Barbosa and Feng [13] or Agarwal et al. [14]. Network Database.SentiBot then examines the profiles of users in . For each user and the network constructed is the subgraph of the Twitter follower network induced by the set of users. There is an edge from user in this subgraph if u follows . In the case of IEDS, there were over 40 million edges considered in total.

Tweet Syntax

We used the following variables to describe the syntax of tweets posted by a user. These are common to many other studies (see, e.g., Yang, Harkreader, and Gu [5]).

1) Average number of hashtags: This is the average number of hashtags used in a tweet posted by user u.
2) Average number of user mentions: This is the average number of other users mentioned by user in his tweets.
3) Average number of links: This is the average number of links present in tweets by user u.
4) Average number of special characters: This is the average number of special characters (e.g., emoticons) present in tweets by user .

IV. METHODOLOGY

Entropy Component

The entropy component detects periodic or regular timing of the messages posted by a Twitter user. On one hand, if the entropy or corrected conditional entropy is low for the intertwist delays, it indicates periodic or regular behaviour, a sign of automation. More specifically, some of the messages are posted via automation, i.e., the user may be a potential bot. On the other hand, high entropy indicates irregularity, a sign of human participation.

Entropy Measures

The entropy rate is a measure of the complexity of a process. The behaviour of bots is often less complex than that of Humans, which can be measured by entropy rate. A low entropy rate indicates a regular process, whereas a high entropy rate indicates a random process. A medium entropy rate indicates a complex process, i.e, a mix of order and disorder .

The entropy rate is defined as either the average entropy per random variable for an infinite sequence or as the conditional entropy of an infinite sequence. Thus, as real data sets are finite, the conditional entropy of finite sequences is often used to estimate the entropy rate. To estimate the entropy rate, we use the corrected conditional entropy.

Spam Detection Component

The spam detection component examines the content of tweets to detect spam. The implementation used for our machine learning component is CRM114 [20]. CRM114 is a powerful text classification system that offers a variety of

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 12, December 2018

different classifiers. The default classifier for CRM114 is Orthogonal Sparse Bigram (OSB), a variant of Bayesian classification, which has been shown to perform well for e-mail spam filtering. OSB differs from other Bayesian classifiers in that it treats pairs of words as features. OSB first chops the whole input into multiple basic units with five consecutive words in each unit. Then, it extracts four word pairs from each unit to construct features, and derives their probabilities. Finally, OSB applies Bayes theorem to compute the overall probability that the text belongs to one class or another Account Properties Component Besides inter tweet delay and tweet content, some Twitter account-related properties are very helpful for the user classification. As shown in Section 3.3, obvious difference exists between the human and bot categories. The first property is the URL ratio. The ratio indicates how often a user includes external URLs in its posted tweets. External URLs appear very often in tweets posted by a bot. The second property is tweeting device makeup, about 70 percent tweets of human are posted via web and mobile devices (referred as manual devices), whereas about 87 percent tweets of bot are posted via API and other auto-piloted programs (referred as auto devices). The third property is the followers to friends ratio.

V. EVALUATION

In this section, we first evaluate the accuracy of our classification system based on the ground-truth set. Then, we apply the system to classify the entire data set of over 500,000 users collected. With the classification results, we further speculate the current composition of Twitter user population. Finally, we discuss the robustness of the proposed classification system against possible evasions

Cross Validation of Accuracy

We use Weka, a machine learning tool, to implement the Random Forest-based classifier. We apply cross validation with 10-folds to train and test the classifier over the ground-truth set. The data set is randomly partitioned into 10 complementary subsets with equal size. In each round, one out of 10 subsets is retained as the test set to validate the classifier, while the remaining nine subsets are used as the training set to train the classifier. At the beginning of a round, the classifier is reset and retrained. Thus, each round is an independent classification procedure, and does not affect subsequent ones. The individual results from 10 rounds are averaged to generate the final estimation. The advantage of cross validation is that, all samples in the data set are used for both training and validation, while each sample is validated exactly once.

VI. CONCLUSION

In this paper, we have studied the problem of automation by bots on Twitter. As a popular web application ,Twitter has become a unique platform for information sharing with a large user base. However, its popularity and very open nature have made Twitter a very tempting target for exploitation by automated programs, i.e., bots. The problem of bots on Twitter is further complicated by the key role that automation plays in everyday Twitter usage. To better understand the role of automation on Twitter, we have measured and characterized the behaviours of humans, bots, on Twitter. By crawling Twitter, we have collected one month of data with over 500,000 Twitter users with more than 40 million tweets. Based on the data, we have identified features that can differentiate humans, bots, on Twitter. Using entropy measures, we have determined that humans have complex timing behaviour, i.e., high entropy, whereas bots are often given away by their regular or periodic timing, i.e., low entropy. In examining the text of tweets, we have observed that a high proportion of bot tweets contain spam content. Lastly,

1) we have discovered that certain account properties, like external URL ratio and tweeting device makeup, are very helpful on detecting automation Bots flip-flop much less frequently than humans in terms of sentiment;

2) When humans express positive sentiment, they tend to express stronger positive sentiment than bots;

3) A similar (but slightly more nuanced) trend holds in terms of expression of negative sentiments by humans; and 4) Humans disagree more with the general sentiment of the application's Twitter population than bots. The endings indicate that engineered features that were previously used to detect fake accounts generated by bots, at best predicted fake accounts generated by humans with an F1 score of 49.75%. This can be attributed to the fact that humans have different characteristics and behaviours than bots which cannot be modelled similarly

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 12, December 2018

REFERENCES

[1] "Top Trending Twitter Topics for 2011 from What the Trend," http://blog.hootsuite.com/top-twitter-trends-2011/, Dec. 2011.
[2] "Amazon Comes to Twitter," http://www.readwriteweb.com/ archives/amazon_comes_to_twitter.php, Dec. 2009.
[3] "Best Buy Goes All Twitter Crazy with @Twelpforce," http:// twitter.com/in_social_media/status/2756927865, Dec. 2009.
[4] "Barack Obama Uses Twitter in 2008 Presidential Campaign," http://twitter.com/BarackObama/, Dec. 2009.
[5] J. Sutton, L. Palen, and I. Shlovski, "Back-Channels on the Front Lines: Emerging Use of Social Media in the 2007 Southern California Wildfires," Proc. Int'l ISCRAM Conf., May 2008.
[6] S. Gianvecchio, M. Xie, Z. Wu, and H. Wang, "Measurement and Classification of Humans and Bots in Internet Chat," Proc. 17th USENIX Security Symp., 2008.
[7] S. Yardi, D. Romero, G. Schoenebeck, and D. Boyd, "Detecting Spam in a Twitter Network," First Monday, vol. 15, no. 1, Jan. 2010.
[8] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," Proc. Seventh ACM SIGCOMM Conf. Internet Measurement, 2007.
[9] S. Wu, J.M. Hofman, W.A. Mason, and D.J. Watts, "Who Says What to Whom on Twitter," Proc. 20th Int'l Conf. World Wide Web, pp. 705-714, 2011.
[10] H. Kwak, C. Lee, H. Park, and S. Moon, "What Is Twitter, a Social Network or a News Media?"Proc. 19th Int'l Conf. World Wide Web, pp. 591-600, 2010.
[11] I.-C.M. Dongwoo Kim, Y. Jo, and A. Oh, "Analysis of Twitter Lists as a Potential Source for Discovering Latent Characteristics of Users," Proc. CHI Workshop Microblogging: What and How Can We Learn From It?, 2010.
[12] D. Zhao and M.B. Rosson, "How and Why People Twitter: The Role that Micro-Blogging Plays in Informal Communication at Work," Proc. ACM Int'l Conf. Supporting Group Work, 2009.
[13] K. Starbird, L. Palen, A. Hughes, and S. Vieweg, "Chatter on the Red: What Hazards Threat Reveals about the Social Life of Microblogged Information," Proc. ACM Conf. Computer Supported Cooperative Work, Feb. 2010.
[14] P. Graham, "A Plan for Spam," http://www.paulgraham.com/ spam.html, Jan. 2008.
[15] J.A. Zdziarski, Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification. No Starch Press, 2005.
[16] M. Xie, H. Yin, and H. Wang, "An Effective DefenseAgainst Email Spam Laundering," Proc. 13th ACM Conf. Computer and Comm. Security, 2006.
[17] J. Yan, "Bot, Cyborg and Automated Turing Test," Proc. 14th Int'l Workshop Security Protocols, Mar. 2006.
[18] Twitter, "Twitter api Wiki," http://apiwiki.twitter.com/, Feb. 2010.
[19] Tweetadder, "Automatic Twitter Software," http://www. tweetadder.com/, Feb. 2010. [39] T.M. Cover and J.A. T
[20] CRM114 "https://info.cs.uab.edu/zhang/Spam-mining-papers/CRM114_Revealed_20050929.pdf"